TITLE:

実験計画法における計画行列の生成について (最尤法とベイズ法)

AUTHOR(S):

田中, 研太郎

CITATION:
田中, 研太郎. 実験計画法における計画行列の生成について (最尤法とベイズ法). 数理解析研究所講究録 2019, 2124: 82-86

ISSUE DATE:
2019-08

URL:
http://hdl.handle.net/2433/252204

RIGHT:

京都大学学術情報リポジトリ
KURENAI 紅
Kyoto University Research Information Repository

数理解析研究所講究録 第2124巻 2019年 82-86

実験計画法における計画行列の生成について

成蹊大学・経済学部・田中研太郎
Kentaro Tanaka
Faculty of Economics,
Seikei University

概要

効率的な実験計画 (なるべく少ない実験回数でなるべく多くの情報が得られる実験計画) を生成する問題が, Group Lasso の形に定式化できることを, やや省略気味で紹介する. また, 実験計画法に計算代数の手法を応用する 2 つのアプローチについて, 例を用いながら解説を行う. これらの解説についてのスライドと成果物であるプログラムについては `https://github.com/tanaken-basis/explasso` において公開している.

1 実験計画の生成について

まず, 効率的な実験計画を生成する問題が, Group Lasso [3] と呼ばれるものの形に定式化できることを紹介していく. 実験計画法と呼ばれる分野で取り上げられる問題には様々なものがあるが, その中の一つとして,「良い実験計画を求める」という問題がある. 「良さ」の基準にも様々なものがあるが, ここでは大雑把に「なるべく少ない実験回数でなるべく多くの情報が得られる」ことを「良さ」の規準として考えることにする.

以下, 簡単のために, -1 か 1 の値のみをとり得る 3 つの変数 x_1, x_2, x_3 があると仮定する. x_1, x_2, x_3 と y の間の関係として, 以下の重回帰モデルを仮定する.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon \tag{1}$$

ここで, β_0 は定数項であり, ϵ は誤差であると考える. そして, $\beta_0, \beta_1, \beta_{2b}, \beta_3$ は未知の偏回帰係数であり, いくつかの実験データを得ることで $\beta_0, \beta_1, \beta_2, \beta_3$ のうちのいくつか (4 つ全部でなくてもよい) の値を推定することを考えたい (実験計画法においてはモデルの全てのパラメーターの値を推定せずに一部だけに興味を持つことはよくあることである).

いま, x_1, x_2, x_3 は -1 か 1 の値のみをとり得るので, 考えられる実験のパターンとしては, 以下の $2^3 = 8$ 通りがある.

実験番号	1	2	3	4	5	6	7	8
(定数)	1	1	1	1	1	1	1	1
x_1	-1	-1	-1	-1	1	1	1	1
x_2	-1	-1	1	1	-1	-1	1	1
x_3	-1	1	-1	1	-1	1	-1	1
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8

(2)

各々の列が, 1 つの実験に対応していることに注意が必要である. 慣例的な計画行列の書き方では各行が 1 つの実験にしているが, 後半の計算代数の説明に合わせるために各列が 1 つの実験に対応する書き方をしている. また, 説明を簡単にするために, 同じパターンの実験は繰り返さないことにしている. 定数 β_0 については $\beta_0 = \beta_0 \cdot 1$ として常に 1 の値のみをとる変数があると仮想的に考える. そして, g 番目の実験で得られた y の値を y_g と表している.

この節の以下においては, これらの実験データから β_1 と β_2 の値のみを推定したいという設定で考える. そして, $y_1, \ldots, y_8$ の線形結合で, それらの値を不偏推定することを考える. さらに, できれば, $y_1, \ldots, y_8$ をす

べて使うのではなく, 一部だけで済むのであれば, その分の実験をしなくて済むのでありがたい. 以下では, $y_1, \ldots, y_8$ の線形不偏推定量とその性質について考えていき, 「なるべく少ない実験回数でなるべく多くの情報が得られる」ようにする方法を構築していく. β_1 と β_2 の線形推定量を以下のように表すことにする.

$$\hat{\beta}_1 = \sum_{g=1}^{8} w_{1g} y_g \quad , \quad \hat{\beta}_2 = \sum_{g=1}^{8} w_{2g} y_g \tag{3}$$

まず, $\boldsymbol{w}_1 = (w_{11}, \ldots, w_{18})^T$, $\boldsymbol{w}_2 = (w_{21}, \ldots, w_{28})^T$ と置き, さらに M を以下のように置く.

$$M = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \end{pmatrix} \tag{4}$$

このとき, $\hat{\beta}_1$ が不偏であることは,

$$M\boldsymbol{w}_1 = (0, 1, 0, 0)^T \tag{5}$$

であることとして表される. 同様に, $\hat{\beta}_2$ が不偏であることは,

$$M\boldsymbol{w}_2 = (0, 0, 1, 0)^T \tag{6}$$

であることとして表される. 次に, 「なるべく多くの情報が得られる」ということを「推定量の分散がなるべく小さい」こととしてとらえることにすると, いくつかの仮定のもとで, $\hat{\beta}_1 = \sum_{g=1}^{8} w_{1g} y_g$ の分散は

$$\sum_{g=1}^{8} w_{1g}^2 \tag{7}$$

となることが計算できる. 同様に $\hat{\beta}_2 = \sum_{g=1}^{8} w_{2g} y_g$ の分散も

$$\sum_{g=1}^{8} w_{2g}^2 \tag{8}$$

となることが計算できる. 最後に, 「なるべく少ない実験回数」を実現することについて考える. いま, 仮に, $\hat{\beta}_1 = \sum_{g=1}^{8} w_{1g} y_g$ と $\hat{\beta}_2 = \sum_{g=1}^{8} w_{2g} y_g$ において, $w_{12} = w_{22} = 0$ であったとする. すると, $\hat{\beta}_1$ においても $\hat{\beta}_2$ においても y_2 の値は使わないということであり, これは, 2 番目の実験が不要であるということになる. つまり, w_{1g} と w_{2g} を同じグループだとみなして, 8 つのグループを作り, $\sqrt{w_{1g}^2 + w_{2g}^2}$ という罰則項を導入した何らかの Group Lasso [3] を考えれば, 「なるべく少ない実験回数」が実現できそうであると分かる. 以上の結果をまとめることで, 「なるべく少ない実験回数でなるべく多くの情報が得られる」ような不偏な線形推定量を構築する問題は, 以下のような制約条件付きの Group Lasso の形の問題として定式化される.

$$\begin{gathered} \min_{\boldsymbol{w}_1, \boldsymbol{w}_2} \ \left(\sum_{g=1}^{8} w_{1g}^2 + \sum_{g=1}^{8} w_{2g}^2\right) + \sum_{g=1}^{8} \lambda_g \sqrt{w_{1g}^2 + w_{2g}^2} \\ \text{s.t.} \quad M\boldsymbol{w}_1 = (0, 1, 0, 0)^T, \ M\boldsymbol{w}_2 = (0, 0, 1, 0)^T \end{gathered} \tag{9}$$

このような定式化においては, チューニングパラメーターの値に注意が必要である. 例えば, もし, チューニングパラメーターを $(\lambda_1, \ldots, \lambda_8) = (1, 1, 1, 1, 1, 1, 1, 1)$ などと, 問題が対称性を持つように設定すると, すべての実験結果が選択されてしまい, 疎な解 (少ない実験回数の解) が得られない. どのように問題の対称性を破るのかが重要であると考えられる.

2 実験計画法と計算代数

計算代数統計は, Pistone and Wynn [2] において計算代数を実験計画法に応用したことが始まりの 1 つだと言われている [1]. まず, Pistone-Wynn 流の計算代数の実験計画への応用方法をおおまかに説明してから, 別の応用方法について説明する. Pistone-Wynn 流の計算代数の実験計画への応用では, たとえば, 以下のような実験計画が与えられているとする.

$$\begin{array}{c|cccc} \text{実験番号} & 1 & 2 & 3 & 4 \\ \hline \text{(定数)} & 1 & 1 & 1 & 1 \\ x_1 & 1 & 1 & -1 & -1 \\ x_2 & 1 & -1 & 1 & -1 \\ x_3 & 1 & -1 & -1 & 1 \\ \hline & y_1 & y_2 & y_3 & y_4 \end{array} \tag{10}$$

これは, 3 つの変数 x_1, x_2, x_3 に対する 4 つの実験からなる計画行列で, とくに, L_4 直行表などと呼ばれる実験計画である. この実験計画 (10) 式において, 例えば, 以下のようなことが分かる.

- x_1 の行を 2 乗したものは, 定数の行と同じである. つまり, この実験では x_1^2 と定数を区別できず, これらの効果を同時には推定できない.
- x_1 の行と x_2 の行をかけたものは, x_3 の行と同じである. つまり, この実験では x_1x_2 と x_3 を区別できず, これらの効果を同時には推定できない.

よって, 計画行列 (10) 式における行同士の関係を見ていくと,

- $x_1^2 - 1 = 0$
- $x_1x_2 - x_3 = 0$

などのような関係式が導かれることが分かる. このような関係式は, 計画行列の関係を行ごとに考えてグレブナー基底を作ることで, 以下のように列挙することができる.

$$x_1^2 - 1 = 0, \quad x_2^2 - 1 = 0, \quad x_3^2 - 1 = 0, \tag{11}$$

$$x_1x_2 - x_3 = 0, x_2x_3 - x_1 = 0, x_1x_3 - x_2 = 0 \tag{12}$$

さらにこれらの多項式から, 以下の標準単項式と呼ばれるものを計算することができる.

$$1,\ x_1,\ x_2,\ x_3 \tag{13}$$

そして, これらの標準単項式の効果だけであれば, 計画行列 (10) 式の実験で推定できる. つまり, Pistone-Wynn 流の計算代数の実験計画への応用では, (10) 式のような計画行列が与えられたときに, この計画行列を行ごとに見て, 計算代数的な手法で導出される標準単項式からなるモデル

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon \tag{14}$$

におけるパラメーター $\beta_0, \beta_1, \beta_{2b}, \beta_3$ の推定が可能であることが分かる. Pistone-Wynn 流の計算代数の実験計画への応用をより短い言葉で言えば, 「与えられた計画行列に対して推定可能なモデルがどのようなものであるのかを明らかにする手法である」と言い表すことができる.

次に, 計画行列を列ごとに見た場合に, 計算代数の手法がどのように応用できそうかを新しく考えてみる. 今度は, まず, モデルが与えられているとする. そして, そのモデルのパラメーターを推定するためにはどのような計画行列が必要なのかを考えていく. まず, 推定したいパラメーターを持つモデルが, 以下のように与えられたとする.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon \tag{15}$$

そして, 以下の全実験 1 から 8 の, どれがあればパラメーターを推定できるのかを考えていく.

$$\begin{array}{c|cccccccc}
\text{実験番号} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
\hline
(\text{定数}) & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
x_1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
x_2 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\
x_3 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
\hline
 & y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8
\end{array} \tag{16}$$

ここで, 計画行列 (16) 式の列ごとの関係を考えてみる. 例えば, 1 と 4 の列を足し, 2 と 3 の列を引くと, ゼロベクトルになる. つまり, 全ての効果が打ち消されて, $y_1 + y_4 - y_2 - y_3$ は平均的にはゼロになるはずである. よって, $y_1 + y_4 - y_2 - y_3$ は, パラメーターの推定においては役には立たない. 次に, 例えば, 全実験の行列において, x_1 の行を消したものを考える.

$$\begin{array}{c|cccccccc}
\text{実験番号} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
\hline
(\text{定数}) & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
x_2 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\
x_3 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
\hline
 & y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8
\end{array} \tag{17}$$

このとき, 例えば, (17) 式の 5 の列から 1 の列を引くとゼロベクトルになる. つまり, x_1 以外の効果が打ち消されて, $y_5 - y_1$ には x_1 の効果のみが現れる. ただし, x_1 の行があるとすると, その成分はゼロにならない. よって, $y_5 - y_1$(の定数倍) によって (1 列と 5 列の実験によって), モデル (15) 式のパラメーター β_1 を推定できることが分かる. ここで, $y_1 + y_4 - y_2 - y_3$ でも (17) 式でゼロベクトルになるが, x_1 の効果も消えてしまうので, β_1 の推定には役には立たないことに注意が必要である. つまり, x_1 の行を消した時だけゼロベクトルになる列の組み合わせで β_1 の不偏推定ができることが分かる. 同様に考えていくと, 全実験からなる計画行列 (16) 式と x_1 の行を消した計画行列 (17) 式において, ゼロベクトルになる列の組み合わせをそれぞれ考えることで, β_1 の任意の線形不偏推定量が以下の多項式の線形結合で表されることが分かる.

$$y_5 - y_1,\ y_2 - y_6,\ y_3 - y_7,\ y_4 - y_8 \tag{18}$$

以上のことをより一般的に考えると, 全実験からなる計画行列とその一部の行を消したものを列ごとに見て, それぞれのトーリックイデアルと呼ばれるものを計算して比較することで, モデルの推定に必要な実験の組み合わせの最小単位が導かれることが分かる. ここで考えた計画行列を列ごとに見る手法について, より短い言葉で言えば, 「与えられたモデルに対して推定に必要な計画行列がどのようなものであるのかを明らかにする手法である」と言い表すことができる. モデルの推定に必要な実験の組み合わせの最小単位が分かれば, (9) 式のチューニングパラメーターの設定において, どのように対称性を破ることができるのかについて, ある程度の指針を与えることができると考えられる.

3　まとめ

効率的な実験計画 (なるべく少ない実験回数でなるべく多くの情報が得られる実験計画) を生成する問題が, Group Lasso [3] の形に定式化できることを紹介した．このような定式化において, 適切な解 (疎な解) を得るには, 対称性をどう破るかが重要である．また, 実験計画法に対する計算代数の応用の 2 通りの方法について解説を行った．1 つは, Pistone-Wynn 流の計算代数の実験計画への応用で「与えられた計画行列に対して推定可能なモデルがどのようなものであるのかを明らかにする手法」であり, もう 1 つは,「与えられたモデルに対して推定に必要な計画行列がどのようなものであるのかを明らかにする手法」である．それぞれ, 計画行列を行ごとに見るのか, それとも, 列ごとに見るのか, という双対的な視点の違いが応用方法の違いにつながっている．これらの解説についてのスライドと成果物であるプログラムについては `https://github.com/tanaken-basis/explasso` において公開している．

参考文献

[1] 青木 敏: 計算代数統計 グレブナー基底と実験計画法, 共立出版, 2018.

[2] Pistone,G and Wynn, H.P: RGeneralised confounding with Gröbner bases, *Biometrika*, **83**(3), (1996), 653-666.

[3] M. Yuan and Y. Lin: Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B (Methodological)*, **68**(1), (2007), 49–67.